

=====
Sequence Listing was accepted.

If you need help call the Patent Electronic Business Center at (866)
217-9197 (toll free).

Reviewer: Mark Spencer

This validation was run special. 05/25/10
=====

Input Set: C:\CRFValidator\10645822.raw

Output Set: C:\CRFValidator\10645822.pdf

Started: 2010-05-25 13:57:56.278

Finished: 2010-05-25 14:31:05.358

Elapsed: 0 hr(s) 33 min(s) 9 sec(s) 80 ms

Total Warnings: 7630

Total Errors: 0

No. of SeqIDs Defined: 80426

Actual SeqID Count: 80426

Error code	Error Description
W 402	Undefined organism found in <213> in SEQ ID (11199)
W 402	Undefined organism found in <213> in SEQ ID (11200)
W 402	Undefined organism found in <213> in SEQ ID (11201)
W 402	Undefined organism found in <213> in SEQ ID (11202)
W 402	Undefined organism found in <213> in SEQ ID (11203)
W 402	Undefined organism found in <213> in SEQ ID (11204)
W 402	Undefined organism found in <213> in SEQ ID (11205)
W 402	Undefined organism found in <213> in SEQ ID (11206)
W 402	Undefined organism found in <213> in SEQ ID (11207)
W 402	Undefined organism found in <213> in SEQ ID (11208)
W 402	Undefined organism found in <213> in SEQ ID (11209)
W 402	Undefined organism found in <213> in SEQ ID (11210)
W 402	Undefined organism found in <213> in SEQ ID (11211)
W 402	Undefined organism found in <213> in SEQ ID (11212)
W 402	Undefined organism found in <213> in SEQ ID (11213)
W 402	Undefined organism found in <213> in SEQ ID (11214)
W 402	Undefined organism found in <213> in SEQ ID (11215)
W 402	Undefined organism found in <213> in SEQ ID (11216)
W 402	Undefined organism found in <213> in SEQ ID (11217)
W 402	Undefined organism found in <213> in SEQ ID (11218)

Input Set: C:\CRFValidator\10645822.raw

Output Set: C:\CRFValidator\10645822.pdf

Started: 2010-05-25 13:57:56.278

Finished: 2010-05-25 14:31:05.358

Elapsed: 0 hr(s) 33 min(s) 9 sec(s) 80 ms

Total Warnings: 7630

Total Errors: 0

No. of SeqIDs Defined: 80426

Actual SeqID Count: 80426

Error code	Error Description
	This error has occurred more than 20 times, will not be displayed
W 213	Artificial or Unknown found in <213> in SEQ ID (80420)
W 213	Artificial or Unknown found in <213> in SEQ ID (80421)
W 213	Artificial or Unknown found in <213> in SEQ ID (80422)
W 213	Artificial or Unknown found in <213> in SEQ ID (80423)
W 213	Artificial or Unknown found in <213> in SEQ ID (80424)
W 213	Artificial or Unknown found in <213> in SEQ ID (80425)
W 213	Artificial or Unknown found in <213> in SEQ ID (80426)



RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/645,822

DATE: 05/25/2010

TIME: 12:07:58

Input Set : C:\CRF4\abss\input\20090331A\10645822.raw

Output Set: C:\CRF4\ABSS\EXPORT\20090331A\J645822.raw

3 <110> APPLICANT: ALEXANDROV, Nickolai et al.
5 <120> TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND
CORRESPONDING POLYPEPTIDES
6 ENCODED THEREBY
8 <130> FILE REFERENCE: 2750-1571P
10 <140> CURRENT APPLICATION NUMBER: US 10/645,822
11 <141> CURRENT FILING DATE: 2003-08-22
13 <160> NUMBER OF SEQ ID NOS: 80426
15 <170> SOFTWARE: PatentIn version 3.2
17 <210> SEQ ID NO: 1
18 <211> LENGTH: 612
19 <212> TYPE: DNA
20 <213> ORGANISM: Arabidopsis thaliana
22 <400> SEQUENCE: 1
23 aatataataa tcgcagggat agacacagga gctcttacca tgatatgggc aatgacagag 60
25 ctcgcaagaa accccgaact aatgaagaaa gttcaaggcg agattagaga ccgctttggc 120
27 aacaataagg agagaatcac caaagaagat ctaaataaag ttcctttctt aaacatggtg 180
29 atcaaggaaa cattcagatt acaccagta gctcctcttc tgctcccaag ggaaacaatg 240
31 actcacatca aggttcaagg ctatgatatt cctccaaaga gacggatctt agtcaacact 300
33 tgtgcgatag gaagagatcc taaactctgg ataaatccgg aagagttaa ccccgagagg 360
35 tttatcaaca accctgtaga ttatagagga caacatttcg agctcttacc gttcgggtct 420
37 ggtcgaagga tatgtcccgg gatgggggtt gggatcacta tagtggagtt gggacttttg 480
39 aacttactat acttcttcga ttggagagca cctgatggga tgacacataa agatatcgat 540
41 acagaagaag ctggtatact tacagtagtc aagaaagtac ctctcaagct tgtcccagtt 600
43 cgagttcagt ga 612
46 <210> SEQ ID NO: 2
47 <211> LENGTH: 203
48 <212> TYPE: PRT
49 <213> ORGANISM: Arabidopsis thaliana
51 <400> SEQUENCE: 2
53 Asn Ile Ile Ile Ala Gly Ile Asp Thr Gly Ala Leu Thr Met Ile Trp
54 1 5 10 15
57 Ala Met Thr Glu Leu Ala Arg Asn Pro Glu Leu Met Lys Lys Val Gln
58 20 25 30
61 Gly Glu Ile Arg Asp Arg Phe Gly Asn Asn Lys Glu Arg Ile Thr Lys
62 35 40 45
65 Glu Asp Leu Asn Lys Val Pro Phe Leu Asn Met Val Ile Lys Glu Thr
66 50 55 60
69 Phe Arg Leu His Pro Val Ala Pro Leu Leu Leu Pro Arg Glu Thr Met
70 65 70 75 80
73 Thr His Ile Lys Val Gln Gly Tyr Asp Ile Pro Pro Lys Arg Arg Ile
74 85 90 95
77 Leu Val Asn Thr Cys Ala Ile Gly Arg Asp Pro Lys Leu Trp Ile Asn
78 100 105 110

RAW SEQUENCE LISTING

DATE: 05/25/2010

PATENT APPLICATION: US/10/645,822

TIME: 12:07:58

Input Set : C:\CRF4\abss\input\20090331A\10645822.raw

Output Set: C:\CRF4\ABSS\EXPORT\20090331A\J645822.raw

81 Pro Glu Glu Phe Asn Pro Glu Arg Phe Ile Asn Asn Pro Val Asp Tyr
82 115 120 125
85 Arg Gly Gln His Phe Glu Leu Leu Pro Phe Gly Ser Gly Arg Arg Ile
86 130 135 140
89 Cys Pro Gly Met Gly Leu Gly Ile Thr Ile Val Glu Leu Gly Leu Leu
90 145 150 155 160
93 Asn Leu Leu Tyr Phe Phe Asp Trp Arg Ala Pro Asp Gly Met Thr His
94 165 170 175
97 Lys Asp Ile Asp Thr Glu Glu Ala Gly Ile Leu Thr Val Val Lys Lys
98 180 185 190
101 Val Pro Leu Lys Leu Val Pro Val Arg Val Gln
102 195 200
105 <210> SEQ ID NO: 3
106 <211> LENGTH: 190
107 <212> TYPE: PRT
108 <213> ORGANISM: Arabidopsis thaliana
110 <400> SEQUENCE: 3
112 Met Ile Trp Ala Met Thr Glu Leu Ala Arg Asn Pro Glu Leu Met Lys
113 1 5 10 15
116 Lys Val Gln Gly Glu Ile Arg Asp Arg Phe Gly Asn Asn Lys Glu Arg
117 20 25 30
120 Ile Thr Lys Glu Asp Leu Asn Lys Val Pro Phe Leu Asn Met Val Ile
121 35 40 45
124 Lys Glu Thr Phe Arg Leu His Pro Val Ala Pro Leu Leu Leu Pro Arg
125 50 55 60
128 Glu Thr Met Thr His Ile Lys Val Gln Gly Tyr Asp Ile Pro Pro Lys
129 65 70 75 80
132 Arg Arg Ile Leu Val Asn Thr Cys Ala Ile Gly Arg Asp Pro Lys Leu
133 85 90 95
136 Trp Ile Asn Pro Glu Glu Phe Asn Pro Glu Arg Phe Ile Asn Asn Pro
137 100 105 110
140 Val Asp Tyr Arg Gly Gln His Phe Glu Leu Leu Pro Phe Gly Ser Gly
141 115 120 125
144 Arg Arg Ile Cys Pro Gly Met Gly Leu Gly Ile Thr Ile Val Glu Leu
145 130 135 140
148 Gly Leu Leu Asn Leu Leu Tyr Phe Phe Asp Trp Arg Ala Pro Asp Gly
149 145 150 155 160
152 Met Thr His Lys Asp Ile Asp Thr Glu Glu Ala Gly Ile Leu Thr Val
153 165 170 175
156 Val Lys Lys Val Pro Leu Lys Leu Val Pro Val Arg Val Gln
157 180 185 190
160 <210> SEQ ID NO: 4
161 <211> LENGTH: 186
162 <212> TYPE: PRT
163 <213> ORGANISM: Arabidopsis thaliana
165 <400> SEQUENCE: 4
167 Met Thr Glu Leu Ala Arg Asn Pro Glu Leu Met Lys Lys Val Gln Gly
168 1 5 10 15
171 Glu Ile Arg Asp Arg Phe Gly Asn Asn Lys Glu Arg Ile Thr Lys Glu

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/645,822

DATE: 05/25/2010

TIME: 12:07:58

Input Set : C:\CRF4\abss\input\20090331A\10645822.raw

Output Set: C:\CRF4\ABSS\EXPORT\20090331A\J645822.raw

```
172          20          25          30
175 Asp Leu Asn Lys Val Pro Phe Leu Asn Met Val Ile Lys Glu Thr Phe
176          35          40          45
179 Arg Leu His Pro Val Ala Pro Leu Leu Leu Pro Arg Glu Thr Met Thr
180          50          55          60
183 His Ile Lys Val Gln Gly Tyr Asp Ile Pro Pro Lys Arg Arg Ile Leu
184 65          70          75          80
187 Val Asn Thr Cys Ala Ile Gly Arg Asp Pro Lys Leu Trp Ile Asn Pro
188          85          90          95
191 Glu Glu Phe Asn Pro Glu Arg Phe Ile Asn Asn Pro Val Asp Tyr Arg
192          100          105          110
195 Gly Gln His Phe Glu Leu Leu Pro Phe Gly Ser Gly Arg Arg Ile Cys
196          115          120          125
199 Pro Gly Met Gly Leu Gly Ile Thr Ile Val Glu Leu Gly Leu Leu Asn
200          130          135          140
203 Leu Leu Tyr Phe Phe Asp Trp Arg Ala Pro Asp Gly Met Thr His Lys
204 145          150          155          160
207 Asp Ile Asp Thr Glu Glu Ala Gly Ile Leu Thr Val Val Lys Lys Val
208          165          170          175
211 Pro Leu Lys Leu Val Pro Val Arg Val Gln
212          180          185
215 <210> SEQ ID NO: 5
216 <211> LENGTH: 1099
217 <212> TYPE: DNA
218 <213> ORGANISM: Arabidopsis thaliana
220 <400> SEQUENCE: 5
221 gttcgagatg ccctataata tatggtgtgg tggctgtagt tctatgatcg ccaaggggtgt 60
223 tcgatttaat gcagaaaaga agcaagttgg caactattat tctacaaaga tatggagctt 120
225 tgcaatgaaa tcaccatgct gtaaaccatga gatagtcatt cagacagatc ctcagaactg 180
227 cgagtatgtc atcactagtg gtgccagaa gaaggttgag gaatatgaag cagaagatgc 240
229 agaaaccatg gagctcactg ctgagcaaga gaagggaaag cttgcagacc ctttttatcg 300
231 tctagagcac caggaagtag atttacagaa gaagaaaagca gctgaaccgc ttttgggtccg 360
233 tcttcagcga gtctcagatg caagacatgc tgatgattac tccctaaaca aagctctacg 420
235 tgcacaactt aggggacaca gaaaacgtgt agcagaggaa gagactgctt caaggaagct 480
237 aggccttgggt ataagactgc ttccaaagat tgaagaagat attaaagctg cctcaaacgt 540
239 gaaattcaag agcaagtttg ataagaatcg aaaggacaaa cgagcactga tccatgcac 600
241 ttccattttt cccgagtcgt cttactcatc gagcaagaaa agaattggagc tggaagcaaa 660
243 aagaaggaaa ataagcgcag catccgcac aagtctattg agaggaggat tcaaagcttc 720
245 atctttatct acaaaccat cagcaagcaa gcccaagggt tcatcagttt ctctaggtat 780
247 tctgcctggg ctcacatgcaa tcagtcttcc taaaggatat acgatcaagg agtttgatac 840
249 tatgaagctc acagcgcagt ttgcggttg gtatggtaac cgtttttggc tgaaattgac 900
251 aaaaatacct agctttgagt tcaccaatca aactgatcgc aacttcaaac gtttctttcg 960
253 gtttggttct gagtattcca aagtatttac gccccaaaa gatttgggag aaaagatgag 1020
255 gaagagtgc gcttatgcag cagccattca ggatgggttt tttccgtctt cttcagtggt 1080
257 atggtctcca gcagcttga 1099
260 <210> SEQ ID NO: 6
261 <211> LENGTH: 365
262 <212> TYPE: PRT
263 <213> ORGANISM: Arabidopsis thaliana
```

RAW SEQUENCE LISTING

DATE: 05/25/2010

PATENT APPLICATION: US/10/645,822

TIME: 12:07:58

Input Set : C:\CRF4\abss\input\20090331A\10645822.raw

Output Set: C:\CRF4\ABSS\EXPORT\20090331A\J645822.raw

265 <400> SEQUENCE: 6

```
267 Phe Glu Met Pro Tyr Asn Ile Trp Cys Gly Gly Cys Ser Ser Met Ile
268 1 5 10 15
271 Ala Lys Gly Val Arg Phe Asn Ala Glu Lys Lys Gln Val Gly Asn Tyr
272 20 25 30
275 Tyr Ser Thr Lys Ile Trp Ser Phe Ala Met Lys Ser Pro Cys Cys Lys
276 35 40 45
279 His Glu Ile Val Ile Gln Thr Asp Pro Gln Asn Cys Glu Tyr Val Ile
280 50 55 60
283 Thr Ser Gly Ala Gln Lys Lys Val Glu Glu Tyr Glu Ala Glu Asp Ala
284 65 70 75 80
287 Glu Thr Met Glu Leu Thr Ala Glu Gln Glu Lys Gly Lys Lys Leu Ala Asp
288 85 90 95
291 Pro Phe Tyr Arg Leu Glu His Gln Glu Val Asp Leu Gln Lys Lys Lys
292 100 105 110
295 Ala Ala Glu Pro Leu Leu Val Arg Leu Gln Arg Val Ser Asp Ala Arg
296 115 120 125
299 His Ala Asp Asp Tyr Ser Leu Asn Lys Ala Leu Arg Ala Gln Leu Arg
300 130 135 140
303 Gly His Arg Lys Arg Val Ala Glu Glu Glu Thr Ala Ser Arg Lys Leu
304 145 150 155 160
307 Gly Leu Gly Ile Arg Leu Leu Pro Lys Ile Glu Glu Asp Ile Lys Ala
308 165 170 175
311 Ala Ser Asn Val Lys Phe Lys Ser Lys Phe Asp Lys Asn Arg Lys Asp
312 180 185 190
315 Lys Arg Ala Leu Ile His Ala Ser Ser Ile Phe Pro Glu Ser Ser Tyr
316 195 200 205
319 Ser Ser Ser Lys Lys Arg Met Glu Leu Glu Ala Lys Arg Arg Lys Ile
320 210 215 220
323 Ser Ala Ala Ser Ala Ser Ser Leu Leu Arg Gly Gly Phe Lys Ala Ser
324 225 230 235 240
327 Ser Leu Ser Thr Asn Pro Ser Ala Ser Lys Pro Lys Val Ser Ser Val
328 245 250 255
331 Ser Leu Gly Ile Leu Pro Gly Leu Met Gln Ile Ser Leu Pro Lys Gly
332 260 265 270
335 Met Thr Ile Lys Glu Phe Asp Thr Met Lys Leu Thr Ala Gln Phe Ala
336 275 280 285
339 Ala Trp Tyr Gly Asn Arg Phe Trp Leu Lys Leu Thr Lys Ile Pro Ser
340 290 295 300
343 Phe Glu Phe Thr Asn Gln Thr Asp Arg Asn Phe Lys Arg Phe Phe Arg
344 305 310 315 320
347 Phe Val Leu Glu Tyr Ser Lys Val Phe Thr Pro Pro Lys Asp Leu Gly
348 325 330 335
351 Glu Lys Met Arg Lys Ser Asp Ala Tyr Ala Ala Ala Ile Gln Asp Gly
352 340 345 350
355 Phe Phe Pro Ser Ser Ser Val Gly Trp Ser Pro Ala Ala
356 355 360 365
359 <210> SEQ ID NO: 7
360 <211> LENGTH: 363
```

RAW SEQUENCE LISTING

DATE: 05/25/2010

PATENT APPLICATION: US/10/645,822

TIME: 12:07:58

Input Set : C:\CRF4\abss\input\20090331A\10645822.raw

Output Set: C:\CRF4\ABSS\EXPORT\20090331A\J645822.raw

361 <212> TYPE: PRT

362 <213> ORGANISM: Arabidopsis thaliana

364 <400> SEQUENCE: 7

```
366 Met Pro Tyr Asn Ile Trp Cys Gly Gly Cys Ser Ser Met Ile Ala Lys
367 1 5 10 15
370 Gly Val Arg Phe Asn Ala Glu Lys Lys Gln Val Gly Asn Tyr Tyr Ser
371 20 25 30
374 Thr Lys Ile Trp Ser Phe Ala Met Lys Ser Pro Cys Cys Lys His Glu
375 35 40 45
378 Ile Val Ile Gln Thr Asp Pro Gln Asn Cys Glu Tyr Val Ile Thr Ser
379 50 55 60
382 Gly Ala Gln Lys Lys Val Glu Glu Tyr Glu Ala Glu Asp Ala Glu Thr
383 65 70 75 80
386 Met Glu Leu Thr Ala Glu Gln Glu Lys Gly Lys Leu Ala Asp Pro Phe
387 85 90 95
390 Tyr Arg Leu Glu His Gln Glu Val Asp Leu Gln Lys Lys Lys Ala Ala
391 100 105 110
394 Glu Pro Leu Leu Val Arg Leu Gln Arg Val Ser Asp Ala Arg His Ala
395 115 120 125
398 Asp Asp Tyr Ser Leu Asn Lys Ala Leu Arg Ala Gln Leu Arg Gly His
399 130 135 140
402 Arg Lys Arg Val Ala Glu Glu Glu Thr Ala Ser Arg Lys Leu Gly Leu
403 145 150 155 160
406 Gly Ile Arg Leu Leu Pro Lys Ile Glu Glu Asp Ile Lys Ala Ala Ser
407 165 170 175
410 Asn Val Lys Phe Lys Ser Lys Phe Asp Lys Asn Arg Lys Asp Lys Arg
411 180 185 190
414 Ala Leu Ile His Ala Ser Ser Ile Phe Pro Glu Ser Ser Tyr Ser Ser
415 195 200 205
418 Ser Lys Lys Arg Met Glu Leu Glu Ala Lys Arg Arg Lys Ile Ser Ala
419 210 215 220
422 Ala Ser Ala Ser Ser Leu Leu Arg Gly Gly Phe Lys Ala Ser Ser Leu
423 225 230 235 240
426 Ser Thr Asn Pro Ser Ala Ser Lys Pro Lys Val Ser Ser Val Ser Leu
427 245 250 255
430 Gly Ile Leu Pro Gly Leu Met Gln Ile Ser Leu Pro Lys Gly Met Thr
431 260 265 270
434 Ile Lys Glu Phe Asp Thr Met Lys Leu Thr Ala Gln Phe Ala Ala Trp
435 275 280 285
438 Tyr Gly Asn Arg Phe Trp Leu Lys Leu Thr Lys Ile Pro Ser Phe Glu
439 290 295 300
442 Phe Thr Asn Gln Thr Asp Arg Asn Phe Lys Arg Phe Phe Arg Phe Val
443 305 310 315 320
446 Leu Glu Tyr Ser Lys Val Phe Thr Pro Pro Lys Asp Leu Gly Glu Lys
447 325 330 335
450 Met Arg Lys Ser Asp Ala Tyr Ala Ala Ala Ile Gln Asp Gly Phe Phe
451 340 345 350
454 Pro Ser Ser Ser Val Gly Trp Ser Pro Ala Ala
455 355 360
```


STATISTICS SUMMARY

PATENT APPLICATION: US/10/645,822

DATE: 05/25/2010

TIME: 12:07:59

Input Set : C:\CRF4\abss\input\20090331A\10645822.raw

Output Set: C:\CRF4\ABSS\EXPORT\20090331A\J645822.raw

Application Serial Number: US/10/645,822

Alpha or Numeric or Xml: Numeric

Application Class:

Application File Date: 08-22-2003

Art Unit:

Software Application: PatentIN3.2

Total Number of Sequences: 80426

Total Nucleotides: 22882596

Total Amino Acids: 20589777

Number of Errors: 0

Number of Warnings: 9861

Number of Corrections: 0

MESSAGE SUMMARY

341 W: 9861 ((46) "n" or "Xaa" used)